Smart Financial Data: Semantic Web technology transforms Big Data into Smart Data

Insurance Data and Analytics Summit 2013
18 April 2013
David Saul, Senior Vice President & Chief Scientist
State Street
Chief Scientist - Catalyst for Innovation

• Act as a catalyst for innovation across State Street and help foster advanced business processes and technology.

• Key focus areas:

  • Provide insight into the best internal and external developments.
  
  • Create new products, services, processes and markets.
  
  • Adapt to respond more quickly and with more flexibility to opportunities.
  
  • Innovate to evolve, revolutionize and even disrupt the status quo to bring forth new and profitable solutions.
The Data Phenomenon

• Big data refers to the sheer Volume, Velocity and Variety of data that is being accumulated by organizations, whether it be structured or unstructured, internal information or data shared by customers and clients.

• Emerging technologies and tools, e.g. cloud computing, are providing the capabilities to support data analysis, processing and storage; however, they’re being stretched thin.

• Traditional ETL and data warehouse approaches are too slow and too inflexible.

• What’s the solution? How do we make data accessible to the business to efficiently and cost-effectively use all the data at its disposal? How do we add Value to the data?
Semantic Databases Create Smart Data

- Derived from the same concept that Tim Berners-Lee used for the world wide web to create hyperlinks.

- Semantic databases technologies are used to add meaning to each piece of data (semantic mapping).

- Ontologies extend semantic mapping to the relationships between the data and allow users to better evaluate and analyze data to create actionable, timely business insights.

- Other semantic technologies include Natural Language Processing (NLP) and Semantic Search.
Semantic Definitions and Standards

- **RDF (Resource Description Framework):** The data modeling language for the Semantic Web. All Semantic Web information is stored and represented in the RDF.
- **SPARQL (SPARQL Protocol and RDF Query Language):** The query language of the Semantic Web. It is specifically designed to query data across various systems.
- **OWL (Web Ontology Language):** The schema language, or knowledge representation (KR) language, of the Semantic Web.

- Finance Industry standards (including LEI) are being developed by:
  - Enterprise Data Management Council (EDM)
  - Object Management Group (OMG)
  - Financial Industry Business Ontology (FIBO)
Key Benefits for the Business & the Industry

- The ability to exchange data across organizations more efficiently and in a standard format will change the way that we communicate with each other and the broader business.

- When data travels along with its meaning, it is easier to do intelligent risk analysis and make critical business decisions in a shorter period of time.

- Reduces the complexity of data analysis and takes it down to a simpler level – gives business users the tools that they can use to, for example, create an ad hoc risk report without having to require programming.

- Overall savings in development and processing time. Ability to start small and grow incrementally.
What the Future Holds

- As industry standards reach maturity and the tools become more widely available, we will begin to see pilot programs expand into a greater implementation of these technologies.

- Semantic technologies will enable the facilitation of real time data flow both throughout the organization and externally.

- The use of semantic technologies as a standard for regulatory reporting, e.g. the Dodd-Frank Act, as proposed by the EDM Council.

- The evolution of the data scientist role – a new career path for database architects.

- Support for enterprise data governance.